DISCUSSION

Tore E. Dalenius, Brown University
and University of Stockholm

The subject of this meeting - survey errors - has been discussed on several previous occasions. The format of this meeting is, however, somewhat unusual: it is built around a research project; more specifically, four of the five papers are a progress report on **the** first phase of that project, while the fifth paper views the problem of survey errors in the context of Federal statistics.

I will structure my discussion as follows. First, I will present "Some General Remarks" which relate to Dr. Horvitz' paper. Next, I will comment in turn on the three project-related papers. Finally, I will comment on Dr. Duncan's paper.

SOME GENERAL REMARKS

While I agree with Dr. Horvitz that "inadequate attention has been given to the role of science in survey design", I think it is important to realize that the role of what Dr. Horvitz calls "art" (I would prefer some such term as "experience") can and should be made part of the role of science. Survey design calls for considering feasible alternatives as inputs into the design procedure. That aspect can clearly be just as scientific as applying some mathematical formula to attain a solution called "optimum".

There are two major weaknesses in this area:

i. Experience is often poorly documented and hence not made accessible on a broad scale;
ii. Terms, concepts, etc. used in reports on survey design are often ambiguous because of the absence of a standardized vocabulary.

The research project under consideration addresses in fact both of these weaknesses. Especially, it aims at developing a taxonomy of survey errors which is compatible with the theory needed to measure and control these errors. The outcome of the project may eventually provide the conceptual and methodological groundwork for a "Survey Design Information System" which - if created - will greatly enhance our endeavors towards "Total Survey Design".

I have four suggestions relating to these aims:

(1) The work on the project should - as far as the forthcoming taxonomy is concerned - be coordinated with similar work by such organizations as the ASA, the IASS, the National Committee on Statistics (the work on "missing data"), the Social Science Research Council, and, of course, the UN Statistical Office. We must avoid adding to the linguistic confusion by having more than one standard!
(2) The scope of the work should include "relevance". What does it help to estimate some parameter with a high degree of accuracy, if the parameter lacks relevance vis-à-vis the users' problems to the solution of which the survey is expected to contribute?
(3) Place primary consideration on those basic concepts in terms of which various survey errors are expressed, for example on "target population", "sampled population", and "frame".
(4) I would welcome a standardization of the symbols we use. If possible, these standards should be compatible with the keyboards of today's typewriters.

THE PAPER BY DR. LESSLER ON "FRAME ERRORS"

This paper, which focuses on surveys of populations of (distinct) "elements" $E_1, E_2, \ldots, E_N$ of some kind, gives support - as I interpret it - to the third of my four suggestions. Thus, before it is possible to create a taxonomy of "frame errors", we must agree upon the meanings of:

i. the target population of elements, $\{E\}_T$
ii. the sampled population of elements, $\{E\}_S$;
iii. the frame F.

Let me state at once my preferred definition of "frame": any material, device, etc. which is used to provide observational access to a population. Clearly, the frame used determines what the sampled population is.

We may distinguish two major sources of errors (corresponding to "coverage" and "content"):

(1) The frame is not congruent with the target population:
   i. some elements in $\{E\}_T$ are not accessible through F; and
   ii. some elements not in $\{E\}_T$ are accessible through F
(2) The frame comprises some "inadequacies" which may cause errors in the estimation of parameters describing $\{E\}_S$. Thus, the link between F and $\{E\}_S$ may be inadequate. Or the auxiliary information available in F may be erroneous.

In passing, I suggest that it may prove helpful to try to keep a clear distinction between "errors" due to the frame and "inadequacies" of the frame; these latter do not necessarily generate errors in the statistics.

Finally, I would like to make three points of minor importance:

(1) In the paper, the author quotes a definition of "undercoverage" stated in U.S. Bureau of the Census (1977), a reference missing in the list of references. This definition does not agree with that given in some previous references from the same agency.

(2) Is "multiplicity" used in the paper in the same sense as it is used by Dr. Sirken in reference no. 16?

(3) Reference no. 5 is now final.

THE PAPER BY DR. KALSBEEK ON NONRESPONSE

Dr. Kalsbeek's paper provides a thought-provoking review of the terminological chaos which is associated with the uses made of such terms as nonresponse, missing data, incomplete data, undercoverage, etc. If we are going to eliminate, or at least significantly reduce, this chaos, it appears inevitable that we address the following two questions.

(1) Should we have "nonresponse" refer solely to elements accessible through the frame, or should elements in $\{E\}_T$ but not accessible through the frame be taken into account? My own preference, for what it is worth, is the first-mentioned alternative.

This view appears to be consistent with the way "nonresponse" is dealt with in the U.S. Bureau of the Census' survey model.

(2) An element accessible through the frame is by design assigned a formal non-zero probability of selection by the sampling design (using that frame). Some such elements may, however, not in fact have such a probability of selection, reflecting the presence of such operational disturbances as "refusals", "not-at-homes", etc.; hence the actual probability of selection may not be identical with the formal one; it may in fact be zero (for steady refusals). Should the definition of nonresponse be based on the "formal" or the "actual" probability of selection? My own preference is the former alternative.

The identification of three general steps in the survey process:

i. location;
ii. solicitation; and
iii. data collection

may prove helpful in endeavors to cope with the nonresponse problem, and the associated formalism (in terms of the entities d_N, d_m, p_N, p_m, Δ_n, π_n, etc.) makes it possible (but not necessarily easy) to see what is common to various authors' views and what is not.

I will make a few comments concerning some of the six principal dimensions discussed in the paper.

I have no quarrel with Dimension 1, per se. I agree with Dr. Kalsbeek that the "deterministic model" represents an oversimplification of reality. This must not, however, be interpreted to mean that such a model necessarily is inferior to a stochastic model. As used in the Hansen-Hurwitz subsampling scheme from 1946, it may in fact provide a powerful, efficient guide to good survey design.

As to Dimension 2, I think it is highly desirable that statisticians pay attention to parameters other than means and totals. Especially, I want to include various measures of association, and also measures of change over time. This last category may be especially important; we too often see cases where it is acknowledged that nonresponse may seriously bias estimates of level but it is somehow assumed, with no supporting evidence, to be harmless with respect to an estimate of change.

As to Dimension 3, it may be worth noting that sometimes we may have reasonably valid information about the direction of the bias. In such cases, we should be able to exploit this information when computing the confidence interval.

Dimension 4 - no comment.

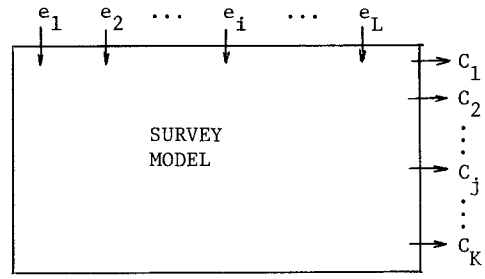
Dimension 5. Here is a case where the choice of definition of nonresponse makes an important difference; the first level is "Ineligible", which includes as a special category "Out of scope for survey observation due to failure to meet eligibility criteria". I would not consider this to constitute nonresponse.

Dimension 6. I would prefer to consider the Hansen-Hurwitz subsampling scheme as an instance of a preventive method rather than a post hoc method.

I suggest adding a Dimension 7, to account for the assessment made of the potential bias due to nonresponse. Some survey-takers seem to think that nonresponse may be a problem in other survey-takers' surveys, but not in their own ...

THE PAPER BY FOLSOM AND LESSLER

The relation of this paper to those on "frame errors" and "nonresponse errors" may be discussed by reference to the following figure, which considers a taxonomy with L classes of survey errors (e_1, e_2, ..., e_L) and K components (C_1, C_2, ..., C_K) of the mean-square error MSE:



where $C_1 + C_2 + \ldots + C_j + \ldots + C_K = MSE$.

The paper gives a brief reference to a "variance decomposition model" developed at the U.S. Bureau of the Census (a proper term would be "mean-square error decomposition model") and a "mixed linear model" from India. These models have in common that the assumption is made that repeat measurements on the same element are independent.

The "Modern Experiment" - influenced by a classic experiment carried out by Karl Pearson - addresses this assumption, in the context of a telephone survey (as distinguished from a nation-wide personal interview survey). While it may be feasible to adopt a design for a state-wide personal interview survey (or for such a survey in a geographically small country), I wonder about its economic feasibility in a nation-wide personal interview survey in the United States.

The section on "Modern Field Studies" states that "most field studies of measurement variance have been developed in the context of simple random sampling". In fact, however, the U.S. Bureau of the Census model has typically been used in the context of "selfweighting samples".

What the authors do is to show how a survey model may be used in the context of "non-self-weighting samples". Their approach suggests two links to classical statistical results; more specifically, to classical variance components models, and to U-statistics. I am not clear as to the validity of this suggestion in the realm of sampling from a finite population without replacement. At any rate, the approach appears interesting and promising, and hence worth our consideration.

THE PAPER BY DUNCAN

The author begins by reviewing past activities of the Office of Federal Statistical Policy and Standards concerning measurement of and reporting of survey errors. Next he points out that "current methodological knowledge" in this area is inadequate; he makes three suggestions for how to improve the situation.

Duncan's conclusion is especially worth noting. Thus he gives in fact his support to endeavors to view the question of errors in surveys in the context of the notion of total survey design. I am especially pleased to quote: "We should accept no statistics which are not accompanied by a carefully developed error statement". It is now up to the Federal agencies to act accordingly.